Building an Intelligent, Automated Tiered Phishing System: Matching the Message Level to User Ability

GIAC (GCIH) Gold Certification

Author: Geoffrey Parker, obiwan324@gmail.com
Advisor: *Dr. Johannes Ullrich*

Accepted: *December 5, 2021*

Abstract

Phishing campaigns and the procedures to run them have remained unchanged since the dawn of the modern era of security awareness training platforms in 2012. The present model uses templates sent at random, assigned based on the level of difficulty of the *template*, not the user. This study creates a new phishing model, method, and process in which the system matches the phishing message difficulty level to the user's skill level. The new design factors the current aptitude of the user and the level of the message difficulty. The system is intelligent, automated, dynamic, and platform-agnostic to scale for the size of the enterprise. Analysis of the tiered system produced statistically significant results indicating that the system improves the user's ability to detect phishing. The system systematically builds the user's skill level and commensurately decreases the risk of falling victim to phishing attacks. There are no other known documented systems like this that are in use at this time. The study discusses forward-looking observations on how practitioners could further enhance this new system when used with User Behavior Analytics and Risk Scoring.

1. Phishing Simulations Have Not Changed in a Decade

Security practitioners are aware that when they engage a commercial phishing system to phish the users, the vendors instruct the practitioners on using the system and conducting phishing assessments (Infosec, 2017). The focus on these procedures is essential to ensure that all (or designated) users encounter phishing attempts, set a schedule for the assessments, capture metrics, and report on the metrics (Cofense, 2021).

The same phishing campaign systems and approaches have now been in use for about a decade, with few changes in the process (Parker, 2019). However, modern platforms are capable of many more techniques for phishing. The traditional phishing method is to create a campaign based on whom to phish, choose the messages to send, choose the frequency and how long to track the campaign, and send (KnowBe4, 2021). The messages are assigned a difficulty level, and the messages are sent *based on the difficulty level of the messages*, not the skill of the users receiving them. Thus, a 0% PPP (Phish Prone Percentage) user could receive a level 1 message as readily as a beginning user receiving a level 5 message. Since this is usually randomized, there is no method or systemic approach to who receives what message. Some organizations pick one message and use that for all persons (Jampen et al. 2020).

Sending the same emails or sending all the emails at the beginning of a campaign can produce the "Bird Dog" or "Prairie Dog" effect (Cherry, 2020). When the user gets the message, they pop up (or engage in chat, email) and say, "Hey, I just got this message, did you? Don't click it. It's phishing". While this organically helps to protect an organization from attacks, in theory, it is not an accurate portrayal of how cybercriminals launch modern phishing attacks in most cases (Verizon, 2021).

While organizations are now mainly working in a remote (Work from Home, WFH) or hybridized model, users still have instant communication tools to alert their team members (Conley, 2020).

1.1. Traditional Campaigns

In many cases, the practitioner's job, usually labeled as a security awareness specialist or trainer, or an IT Security analyst, is to assure the launch of the assessment. The practitioners ensure that valid data is collected, and they provide metrics to a Director, CISO, CIO, or possibly senior management (Spitzner, L., & DeBeaubien, D. (2021).

The industry, for the most part, has adopted a "one size fits all" approach to phishing, sometimes referred to as "throwing spaghetti against the wall" or "Spray and Pray" (Cherry, 2020).

Typical instruction from the commercial platform vendors themselves (KnowBe4, 2021) consists of:

1. Choose who to phish
2. Choose what messages to send
3. Choose how often to phish and how long the campaign runs
4. Launch campaigns
5. Analyze results
6. Add clickers to training groups
7. Report on metrics

The industry has widely adopted a standard 1-5 level of difficulty rating for message templates (with 1 being the lowest and 5 the highest) used in phishing campaigns (KnowBe4, 2021). Commercial phishing platforms commonly use this system of message difficulty rating. Thus, a practitioner would launch a phishing campaign with a random set of messages at all five levels in many cases. There is no strategy for receiving a particular difficulty level/rating message. An experienced user may receive a level 1 message, and a beginning user may receive a level 5 message using this randomized phishing simulation approach.

2. Creating a New Model

The intelligent, automated, tiered phishing model shifts the paradigm. The new model focuses on the user's skill level, assessed by their "Phish Prone Percentage" (PPP) or aggregate failure rate (KnowBe4, 2021). The PPP is not a simple click rate. It is a score based upon the message click rate, open attachment rate, macro-enabled rate, message replied to rate, data entry rate, and whether the user reported the phish. A simple click/attach open/macro enable/data entry rate would suffice for systems that do not have access to all these metrics or for organizations that do not use them.

Type of phishing failure
1. Clicking the embedded link
2. Entering data on a landing page
3. Opening an attachment
4. Enabling a macro on an attachment
5. Replying to the simulated phishing email
6. Report as phishing

Figure 1. Elements of the Phish Prone Percentage (PPP) (Knowbe4, 2021)

By phishing the users at their current skill level, they receive messages designed to fit their experience level. The users can successfully determine if the messages are phishing and report them. This approach gives the user more helpful practice and learning than a randomized method which could send an easy message or a message above the user's skill level.

The system then assigns a PPP and dynamically moves a user into the appropriate tier. The tiered system gives the user more practical experience, similar to how athletes build their skills through practice. The users can develop their skills as well. It can also

generate a more realistic phishing score, which means a more realistic risk model for the organization.

The tiered system bypasses the situation where a beginning level user may receive a very high difficulty level such as BEC / BIC (Business Email Compromise / Business Identity Compromise), which they will not detect as phishing. Conversely, a very advanced 0% PPP user may receive an email at work threatening to close their Facebook account if they don't click the link within 24 hours.

The tiers are created based on the user's PPP. The system can assign users to groups and move them between groups (tiers) using dynamic groups and rules as their score changes. The phishing platform establishes a baseline for new users, and then the user is moved into the appropriate tier.

The tiers are established based on PPP from 0% to 90+ %, and the practitioner assigns message difficulties from 1 to 5. The practitioner can also assign custom message templates to the tiers. Once the system is set up, tested, and functioning well, the administrator can clone the campaigns to run again at the next interval, with chosen messages in the campaigns. The system automatically changes the users per tier based on their most recent phish-prone score and time working at the organization.

2.1. Potential benefits of tiered phishing

If the organization is basing a threat/risk model for phishing compromise on the current PPP or click rate, the model may not be accurate. In the test data for this study, the CIRT (Computer Incident Response Team) declared an incident after users fell victim to an EMOTET attack. Upon review, the risk model indicated a low threat level due to phishing compromise based on the current click rate from phishing assessment campaigns. However, the actual level of compromise was much higher and was off by more than 20%. When The Risk Team reviewed the risk model again from a compromise with the tiered phishing system in use, the risk model was far more accurate, to within about .5%. The tiered system also provides the following potential benefits:

- May provide more accurate threat modeling for the actual phishing attacks
- As the skill level increases, the message difficulty level increases

- Training users to recognize phishes at their skill level allows them to be successful and build phishing recognition skills continuously
- Reduces risk by training users to detect phishing attacks more accurately
- Provides granular risk reporting based on user level of capability
- Provides indicators for where further development and training is needed

2.2. Research Method

The study organization first developed a design requirement to create the intelligent, automated, tiered phishing system. The study organization designed the system to place users into dynamic groups automatically, and users would automatically move in the system based on their current Phish Prone Percentage (PPP).

The system works based on a combination of dynamic groups and rules. The groups are the tiers based on user PPP. The rules dynamically move the users between groups.

The Tiers design was robust in that most users were in the top tiers from 0 to 10% PPP. The initial plan was for two tiers encompassing the 0% PPP group and the 1-10% group. However, this was not granular enough, as most users were in the top two tiers, and the subsequent plan included 1-5% and 6-10% tiers. These top three tiers accounted

Tiered Phishing

We currently have 12 levels or tiers for phishing

% PP = Phish Prone Percentage

Notice that 7,596 of our 8,367 are in the top 2 advanced groups

Tiers are dynamic and users move from group to group as their PP % changes

Enterprise PP % 3.1

Group	User Count	Phishing Level
Most Advanced users: (0.00) % PP	Users: 5425	5
Not as Advance (0-10%)	Users: 2195	4
Greater than Intermediate (11-20%)	Users: 392	3
Kind of Intermediate (21-30%)	Users: 82	3
Intermediate users (31-40%)	Users: 42	3
Less than Intermediate users (41-50%)	Users: 67	3
Best of Beginners (51-60%)	Users: 1	1/2
Better but not best of Beginners (61-70%)	Users: 1	1/2
Beginner Users (71-80%)	Users: 0	1/2
Bad Beginners (81-90%)	Users: 0	1
High Risk Beginners (91%+)	Users: 3	1
New Hires	User: 181	1 for 4 weeks, then based on PP %

Figure 2. Initial Tiered phishing Design

for over 80% of the users in the test. The administrator then created tiers for intermediate and beginner users as the program matured.

Tiered phishing allows us to match phishing simulations to the skill level of the user. This gives us a more accurate risk score for the individual.
This also allows users to continue to improve their skills in detecting phishing attacks

Tiered Phishing

We currently have 12 levels or tiers for phishing

% PP = Phish Prone Percentage

Notice that 7,306 of our 7,887 users are in the top 3 advanced groups

Tiers are dynamic and users move from group to group automatically as their PP % changes

Enterprise PP % currently	2.0
Industry Benchmark PP %	5.1

Group	User Count	Phishing Level (difficulty)
Most Advanced users: (0.00%) PP (never clicked/opened)	Users: 4737	5
Not as Advanced (0-5%)	Users: 1699	4 - 5
Not as Advanced (5-10%)	Users: 975	4
Greater than Intermediate (11-20%)	Users: 386	3
Kind of Intermediate (21-30%)	Users: 72	3
Intermediate users (31-40%)	Users: 25	3
Less than Intermediate users (41-50%)	Users: 3	3
Best of Beginners (51-60%)	Users: 0	1 – 2
Better but not best of Beginners (61-70%)	Users: 0	1 – 2
Beginner Users (71-80%)	Users: 0	1 – 2
Bad Beginners (81-90%)	Users: 0	1 – 2
High Risk Beginners (91%+)	Users: 3	1 – 2
New Hires	Users: 67	1 for 4 weeks, then based on PP %

Figure 3. Finalized Tier Structure as the program matured

2.3. Dynamic Groups and Rules are the basis of tiers

The administrator forms the group containers (tiers) for the users from two elements. The PPP and the rules govern the groups (see Figures 2 and 3 above). A traditional phishing system uses logical groups instead. These groups are static and are updated either through Active Directory integration (or equivalent) or manually. For example, the traditional system organizes groups by business line, department, or possibly by offices in geographic locations. These groups are logical containers for creating phishing campaigns.

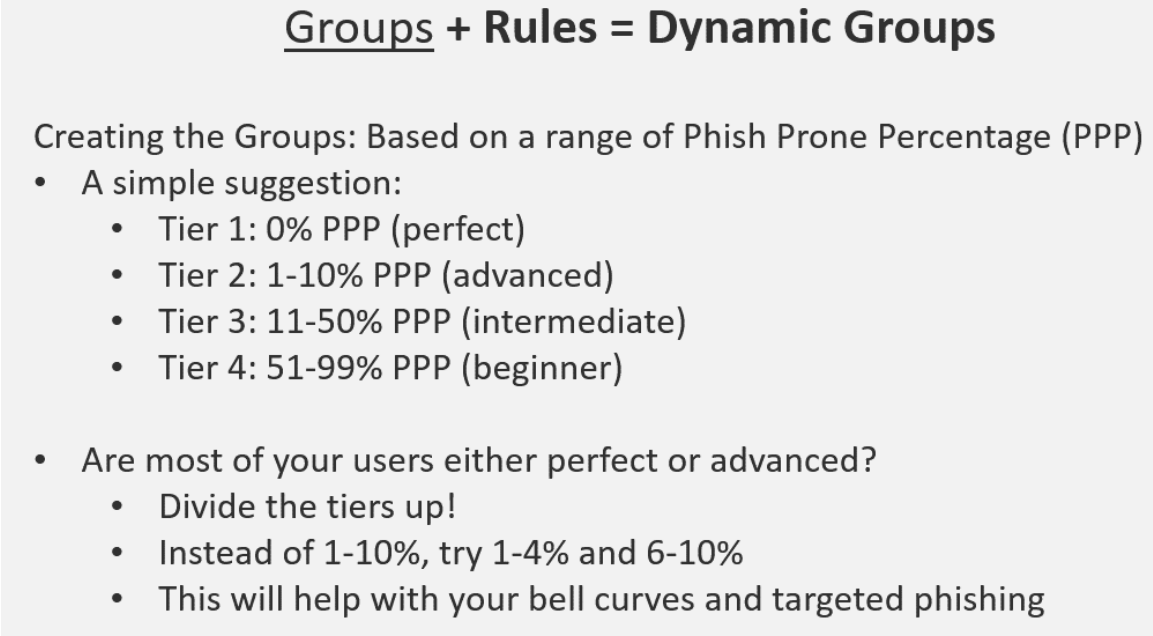


Figure 4. Sample of Groups to create for tiered phishing

Rules dynamically move the users between groups. Two types of rules are required: (1) the rules for the range of Phish Prone Percentage (PPP) and (2) the rules for the date the user began working at the organization. The date rule comes into play again when a baseline is established for new users to the organization.

In Figure 5 below, an example of the rules for the 1-5% PPP group shows in the first rule that the range of PPP must be greater than 0, which demonstrates that this is not a 0% PPP group. The second rule indicates that the highest % of PPP must be less than six. Therefore, this group, by definition, is the 1-5% PPP group. The last rule indicates that the user must have been at the organization for more than 30 days. This rule ensures that the user is not new and has sufficient PPP history to derive a PPP percentage score.

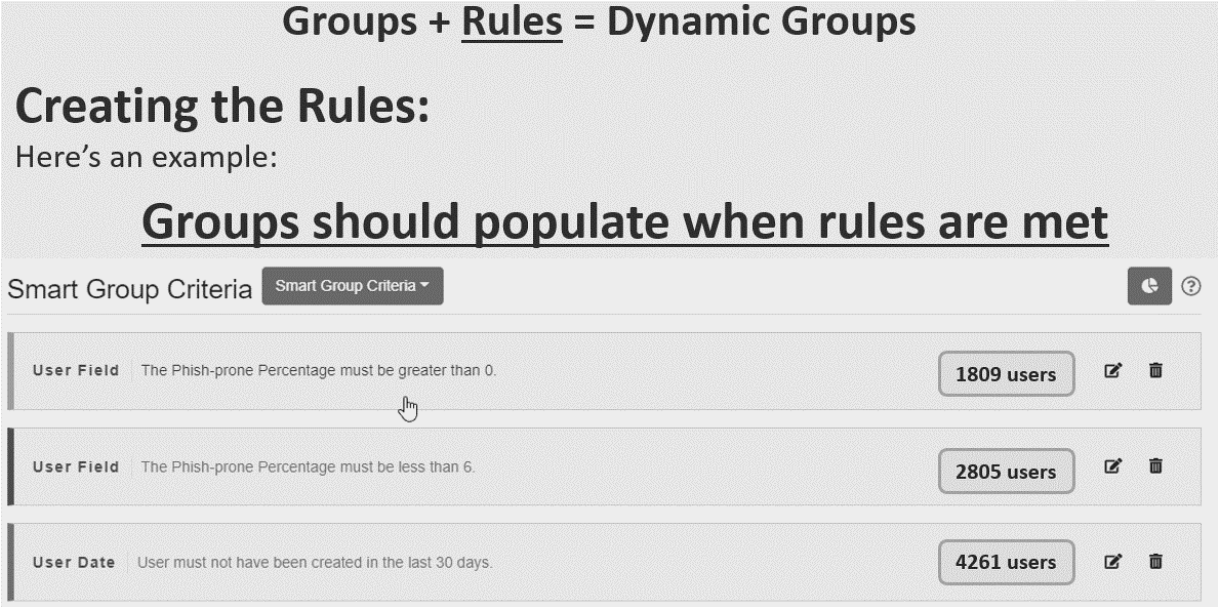


Figure 5. Sample of rules used to create the 1-5% PPP group for tiered phishing

2.4. Creating a baseline and PPP for new users

When new users enter the organization, they have no phishing history and no PPP. The system creates a baseline to address this. In the study, the administrator uses the phishing platform to create an automated phishing campaign that phishes all new users – those who have been at the organization for less than 30 days – one time per week for four weeks. The system uses level 1-2 (very easy to easy) phishing message templates in various administrator-selected categories. Since this is an automated campaign, no administration is required. Once the user has established a > 30-day history, the standard dynamic group rules created for the tiers activate (rule: User must not have been at the company for less than 30 days) and automatically move the user to the proper phishing tier.

The user's experience level is unknown when they enter the organization. Using lower difficulty level templates is more beneficial. The users may pass all "new user" phishing campaigns and are elevated to a higher phishing tier and difficulty level. In that case, when they fail the first few campaigns, the system will intelligently and automatically move them to the correct level.

2.5. Creating and cloning phishing campaigns

Most commercial phishing simulation platforms provide a method to "clone" reoccurring phishing campaigns. Using a "clone" procedure, the administrator can select "clone," then change the name of the campaign, and any other chosen settings, such as "date to run" and other options.

The practitioner can run new periodic campaigns (e.g., monthly) with selected options by cloning the campaign. For example, the practitioner can change landing pages, including attachments, and select message template categories.

The simplest method used in the present experiment is to change the monthly name of the campaign to reflect the current period, the run dates, and the phishing domains, all of which the practitioner can do in about one to two minutes. Another option would be to set the campaigns to run automatically and periodically, just like the beginner campaign. However, since this is an experimental system, and it makes sense to audit it and check performance, the present study opted for monthly cloned campaigns. All modern commercial phishing platforms observed have the option of cloning. Please note that non-commercial platforms, such as the freeware Go-Phish, do not have many of the features referenced herein, for example, dynamic or smart groups and robust rule sets.

Figure 6 shows a sample setup for monthly phishing campaigns using the intelligent, automated tiered phishing system. Notice that the names include both the level and Phish Prone Percentage (PPP) of target users.

There is a fundamental difference between traditional phishing campaigns and the tiered system. In the conventional method of campaigns, practitioners phish users by a logical unit, such as department or geographic location. The *tiered* system phishes users based on *PPP*. Thus, the campaigns are organized by tier using the tiered method.

A possible console view of your campaigns

Campaign	Target Group	Tier	Message Difficulty	User Count	Action
Advanced 0%	0% PPP	1	Very Hard (5 star) ★★★★★	2832	Clone
Very Good 1-10%	1-10% PPP	2	Hard (4 star) ★★★★	2300	Clone
Intermediate 11-50%	11-50% PPP	3	Medium hard (2-3 star) ★★★	520	Clone
Beginner 51-99%	51-90+% PPP	4	Easy (1 star) ★	430	Clone

Figure 6. A sample console view of tiered phishing campaigns

The table in Figure 6 identifies the target group or groups, as it may be best to group several tiers into an "intermediate" campaign. In some organizations, the practitioners find 75-90% of their users in the top two to three tiers. For reporting metrics, since the practitioner may have 5000 users in the top two tiers and only 500 in *all* the intermediate tiers, it is much more sensible to group all the tiers between 11 and 50% (e.g., four tiers of 10 percent increase in PPP each). The practitioner then creates one intermediate campaign for all 500 users labeled intermediate. Since the phishing platform allows the practitioner to select the groups to phish, it is far more expedient to create an "Intermediate" campaign and place all the tiers that fall within the intermediate range into the campaign. The practitioner then performs the same actions with the beginner-level campaigns.

The tier number provides a reference only. The message difficulty sets what message templates at what level (tier) to use. The user count indicates how many users are in the group. Finally, the campaigns have action options (e.g., disable, delete, clone). If the practitioner selects “clone," as detailed in the preceding section, the campaign administrator can take a few minutes to set up all the required monthly campaigns.

2.5.1. Best practices when creating groups

The practitioner must take care of the exact and specific percentage rules to create the groups. Otherwise, a percentage overlap could occur. An example is if the practitioner establishes a group at 1-5% and a second group at 5-10%. Because the design is percentage-driven as controlled by the rules, this would create a scenario where a user would appear in two groups. Instead, a best practice is to create a group where the first rule is PPP is > 0, and the second rule is PPP is < 6. These two rules encompass the 1-5% PPP users. The next group would have the first rule set to PPP is >/= 6, with the second rule being PPP is < 11. These rules create 6-10% PPP users. The experiment identified that using decimals does not work correctly in the rules. There are inevitably multiple users with a score of, e.g., 5.8% PPP, causing these users to exist in both groups. When the group building is complete, it is best to review the users and ensure they do not appear in multiple tiered groups before launching phishing campaigns, so users are not phished multiple times per campaign run.

2.5.2. Best practices when creating campaigns

In the organization where the researcher performed the study and experiments, specific policies do not allow phishing campaigns to address sensitive topics. These range from not threatening the user's job, livelihood, pay, benefits, and a few other sensitive topics. The organization *does* allow departments such as HR and IT to be spoofed, and even user's managers. It is a brief exercise for a campaign administrator to check the phishing platform console and review all new templates added after the previous campaign. A review avoids the potential for missing a newly added system template that would interfere with organizational policies.

Some organizations are also crafting their templates or using benign examples of actual phishing attacks that have recently taken place. This study used approximately 2800 templates across many categories and five difficulty levels per month. The study also scheduled the phishing campaign messages to send over several business days so that messages arrive over time instead of being sent at the beginning of the campaign. Scheduled sending helped prevent the organic or "prairie dog" effects of users casting

about asking if their colleagues had received a phishing message – since the answer was almost universally no, they had not.

3. Findings and Discussion

It is evident for this new model that the study compares the traditional system and the intelligent, automated tiered system to gain a clear picture of the differences and the potential benefits – or lack thereof – from the creation and implementation of such a system. The study gathers 18 months of data before using the tiered phishing system and 18 months of data while exclusively using the tiered system.

The data was captured from a production system and sanitized appropriately to render it anonymous.

3.1. Pre- and post-tiered system comparison

Figure 7 shows a view of the enterprise click rate immediately before implementing the tiered phishing system.

Rank	Click Rates 2019 YTD prior to tiers	Average Click
1	Compliance	0.9990%
2	Financial	1.3825%
3	Insurance	1.7184%
4	Engineering	2.0506%
5	Travel	2.0833%
6	Product Management	2.0888%
7	Information Distribution	2.2831%
8	Product Operations	2.4067%
9	User Administration	2.4323%
10	Management	2.4544%
11	Human Resources	2.4977%
12	Product Group 1	2.7020%
13	Information Technology	2.7757%
14	Product Administration	2.8250%
15	Corporate	2.8571%
16	Legal	2.8686%
17	Financial	2.8750%
18	Financial	2.9502%
19	Communications	2.9529%
20	Commercial products	3.1045%
21	Product Group 2	3.1369%
22	Commercial Products 2	3.2926%
23	Commercial Products 3	3.3286%
24	Products - Government	3.3580%
25	Compliance	3.6012%
26	Property	3.8424%
27	Distribution network	3.8872%
28	Sales	3.8938%
29	Policy and Governance	4.1667%
30	Commercial Products 4	4.6054%

Figure 7. Enterprise click rates before implementation of Tiered phishing system

Notice that one group has a < 1.0 % click rate, and two groups have a > 4.0% click rate. Comparing these click rates with the same departments using the tiered phishing system (Figure 8), the click rate is significantly lower for nearly the complete term of 2021. Also, notice that based on data in Figures 2 and 3, the click rate dropped from a 3.1% average to 2.0% after the organization used the tiered phishing system for a year. Reviewing the detail for Figure 8 the tiered system in use shows three departments at 0% click rate, 1 with < 1.0% click rate, and 0 departments with 4.0% or above. The data shows a statistically significant overall improvement in phishing click rates over the same

duration of time when using messages targeted to the user's ability level compared to messages sent at random. The Significance, with a P=.05, is 0.000524.

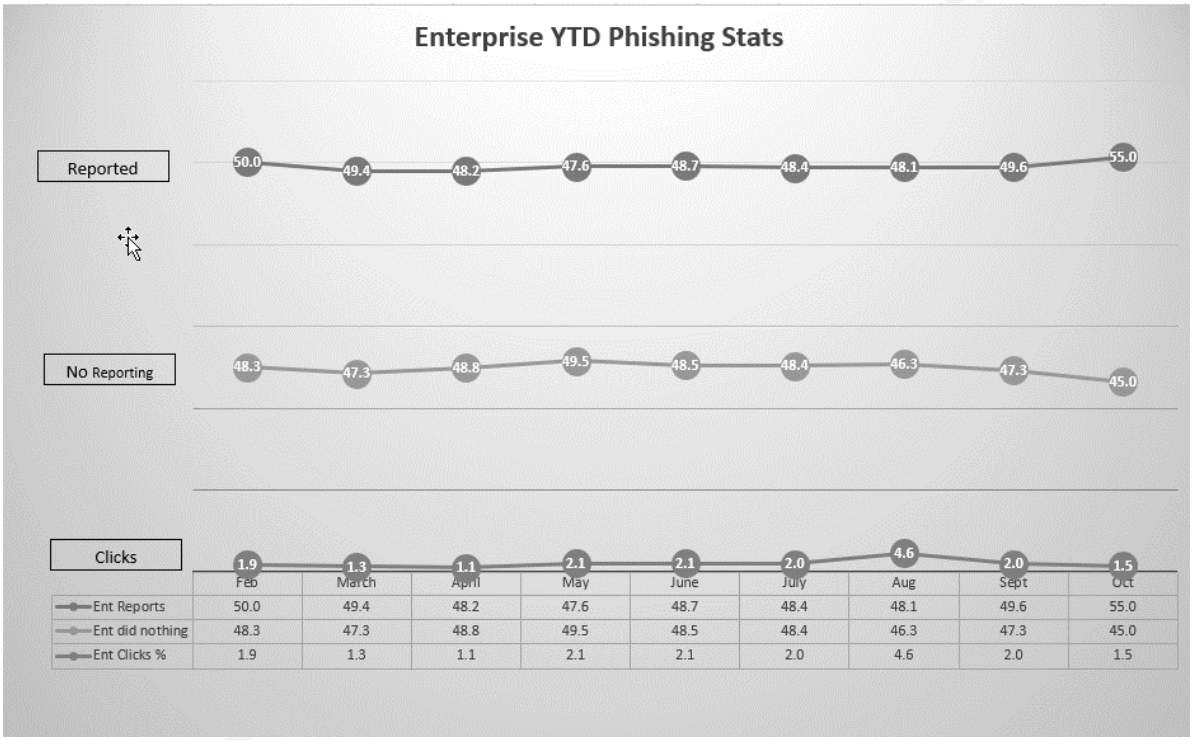
Rank	YTD click rate TIERED 2021	YTD Click %
1	Corporate	0.00%
2	Policy and Governance	0.00%
3	Travel	0.00%
4	Insurance	0.62%
5	Compliance	1.22%
6	Product Management	1.41%
7	Communications	1.66%
8	Commercial Products 2	1.69%
9	Commercial Products 3	1.76%
10	Financial	1.98%
11	Information Distribution	1.99%
12	Human Resources	2.08%
13	Product Group 1	2.03%
14	Commercial Products 4	2.09%
15	Information Technology	2.00%
16	Engineering	2.02%
17	Management	2.10%
18	Product Group 2	2.22%
19	Financial	2.25%
20	Product Operations	2.29%
21	Product Administration	2.31%
22	User Administration	2.38%
23	Distribution network	2.50%
24	Commercial products	2.61%
25	Sales	2.64%
26	Products - Government	2.76%
27	Legal	2.96%
28	Property	3.38%
29	Compliance	3.41%
30	Financial	3.98%

Figure 8. Enterprise click rates using Tiered phishing system

3.2. The metrics with a tiered system

Overall, as the intelligent, automated tiered phishing system matured over time, it became clear that the enterprise results were steadily improving. In addition to the click rate dropping, the phish reporting rate increased, and the "did nothing rate" (no failure, no

report), which was also analyzed, dropped. Figure 9 below reveals the ongoing trend using the tiered phishing system throughout 2021.



	Feb	March	April	May	June	July	Aug	Sept	Oct
Ent Reports	50.0	49.4	48.2	47.6	48.7	48.4	48.1	49.6	55.0
Ent did nothing	48.3	47.3	48.8	49.5	48.5	48.4	46.3	47.3	45.0
Ent Clicks %	1.9	1.3	1.1	2.1	2.1	2.0	4.6	2.0	1.5

Figure 9. Enterprise click rates before implementation of Tiered phishing system

Of significance in this chart are the three metrics observed:

1. Over the year, the reporting rate stayed relatively steady and then began to climb, starting at 50% and rising to 55%, a 5% gain, which shows statistical significance.
2. The No Reporting rate started at 48.3%, dropped to 45.0%, dropping by 3.3% in those who neither failed nor reported. This rate tends to fluctuate due to many variables.
3. The Click rate itself started at 1.9%, and at the last report in October 2021, campaigns had dropped to 1.5%. With an industry average of 5.1% in the study category, this demonstrates a significant improvement and a low failure rate overall. That is very important, particularly when the risk and threat

modeling for the organization (before tiered phishing) was off by a significant margin following an actual, not simulated attack. The organization's confidence in the risk model number is far higher, as the attack simulations are more accurate, as are the user reactions based on the tiered campaigns.

Note the spike in click rate for August 2021. The organization used higher difficulty messages across the board and very challenging and realistic attachments for these campaigns. The organization did this to test the user's proficiency. Based on this campaign data, the organization then implemented a bi-monthly schedule of elevated difficulty level messages with attachments.

3.3. Group (tier) trends of the tiered system

One of the other expected outcomes from utilizing the tiered phishing system is a steady improvement in users' ability to spot phishing attempts. As the users improve, their PPP commensurately trends down, and they continuously move to higher-level groups over time (those with a higher phishing proficiency and a lower PPP).

An example of this has manifested itself in the study group demographics. Over the past five phishing campaigns, there have been *no* users in the beginner groups (Figure 10). When the administrator launches the phishing campaigns, they error out, as there are no users in the groups to phish.

Tier	Month	Number of Users
Beginner	**November**	0
Beginner	**October**	0
Beginner	**September**	0
Beginner	**August**	0
Beginner	**July**	0
Beginner	**June**	9

Figure 10. No users in the Beginner tiers for the past five months

3.4. Anecdotal evidence of tiered phishing system working

Other anecdotal evidence indicates that the tiered system is working and making a positive difference to improve users' abilities to detect, deter, and report phishing.

Before using the tiered phishing system, the administrator responsible for launching and administering the campaigns received little to no feedback regarding the phishing campaigns. Feedback received was usually some form of negative feedback or a complaint, along with protests that the system was wrong, and the user did not actually fail. The administrator responded with the data from the system showing the details of the phishing failure to the user. The caution was that the bad guys would not give the users a break. Therefore, the system structured the campaigns to be as realistic as possible.

Significantly, once the organization implemented the tiered phishing system, the administrator received a steady flow of *positive* comments regarding the campaigns. These comments ranged from chagrin at being caught to comments on how realistic the simulated phish was. Some users commented on how appropriate the phish was, or they felt like they were now getting messages relevant to their skill level. The SOC (Security

Operations Center) personnel have been particularly verbose and positive about the campaigns as evidently; they relish the challenge. Before the organization implemented tiered phishing, they were bored by easily detecting messages. The feedback was collected and reported to management since it served as anecdotal evidence that the program was effective.

3.5. Granular Metrics with tiered phishing

It is possible to gather more metrics than with a traditional system using the tiered phishing system. Most modern commercial phishing platforms or security awareness training systems have robust reporting. However, the following metrics are specific to the tiered phishing system:

- The type of campaign that will always result in failure
- The tier the user is in – difficulty level of message
- The history or group trajectory of user
- The user dwell time in group
- The type of failure (if measured/recorded)
- The number of failures in the last 12 months – at what tier
- The number of users per group (shows user skill level)

These metrics may be helpful and attractive to the organization. The usefulness also depends on the use of the data. An example is a policy that dictates the delivery of mandatory training after *X* number of failures in *Y* months. Another example might be if the user is dropping steadily through tiers instead of staying stable or climbing, indicating another issue or the need for more training. From a broader perspective, the data generated by the tiered phishing system may serve as input for other analyses and reporting, which will be discussed in the next section.

One excellent use of the tiered phishing data is visual analytics. Utilizing Tableau, the administrator of this system and study created a visual analysis of the data accessed using RBAC (Role-Based Access Control). That means that a given manager/director/VP can log in to the Tableau dashboard and see precisely where their department stands compared to the enterprise. They can view the data on specific cost centers and users with

the click of a button. The visual format makes it very easy to understand meaningful data at a glance.

4. Recommendations and Best Practices for Implementing a Tiered Phishing System

Eight foundational principles can help practitioners implement an intelligent, automated tiered phishing system. One of the most important first steps is to contact the vendor of the phishing platform and advise them of what the practitioner desires to create. The CSR (Customer Service Representative) may not know if their system is capable of the desired outcome. However, two key questions answer whether the platform can create a tiered phishing system: 1. Does the platform have smart or dynamic groups? 2. Does the platform have rules to apply to the dynamic groups? The CSR engages technical support as needed.

If the answer to these two questions is yes, the platform can host a tiered phishing system.

Figure 11 below shows eight foundations or best practices the researcher used to build the tiered phishing system. The researcher based this guidance on their experience setting up and operating the *tiered production* phishing system used in this study for over 18 months.

Figure 11. Eight foundations of building a tiered phishing system

Appendix A references a worksheet to allow the prospective tiered phishing system operator to create a system. Appendix B contains other cautionary best practices for creating and implementing a tiered system. One of the most critical best practices *is to*

run In Parallel while developing the tiered phishing system. This allows the practitioner to watch the group/tier flow, fine-tune groups and campaign tiers while running the traditional phishing campaigns. This study observed the flows and adjusted for three months while maintaining the conventional campaign structure before switching to the tiered system.

When the organization switched to the tiered system, there was a sudden jump in the click rate for the organization, owing to more accurate difficulty level messages delivered to more users. That returned to a more usual baseline as the organization utilized the tiered system.

These three cautions or best practices are helpful to the prospective operator of a tiered phishing system.

- If the system uses many tiers, the practitioner may be doing many campaigns— or receiving lopsided metrics – unless the practitioner groups the tiers into campaigns
    - Tiers (groups) can be moved to different campaigns
- Balance against the levels of difficulty for phishing templates
    - Most platforms have five standard levels of difficulty
    - Practitioners can still create their own message templates
- Be careful to increment the PPP when making rules:
    - If there is overlap (e.g., 1-5%, 5-10% instead of 6-10%), the user ends up in 2 groups
    - Do not use decimals for PPP; use whole numbers only

4.1. Implications for Future Research

The intelligent, automated tiered phishing system should be considered a beginning, not an end. Many possibilities for how prospective practitioners can use the system, modify it, or enhance the system are beyond the scope of this study. Different fields controlling different rules, regulations, policies, and procedures may develop new capabilities and directions for the tiered system.

In addition to the uses, functions, and metrics discussed herein, two other future projects are already in play at the organization that hosted this study. The tiered phishing system provides more accurate and detailed elements for both systems under development.

The first example is the development of User Behavior Analytics (UBA). UBA uses a matrix of behaviors, actions, and capabilities. The tiered phishing scores are a vital indicator for the UBA matrix. This complex system grabs data from many areas and compiles it into a visual dashboard that indicates the granular level of risk to the organization.

The second project is the development of a risk score. Such a score relies on inputs of actions and behaviors, which derive from the UBA score. The scoring from the tiered phishing system is a crucial component of such a score. The risk score concept resembles a credit score to the user and indicates the relative level of risk for the given user and area of the organization. When the score is good, it means low risk. When the score is poor, it helps indicate where targeted training or other developmental or remedial actions are needed.

These are but two of the many possible directions that the tiered phishing system can go for future development.

5. Conclusion

The industry has conducted phishing assessments in essentially the same way for about a decade. But the world has moved on since then. New generations of computing, hyper speeds, and modern users who use modern technologies are new elements to consider and factor into phishing assessments. The TTPs (Tools, Tactics, and Procedures) of a burgeoning cybercrime environment with incredibly sophisticated systems and fully educated professionals at the helm have considerably increased the threat vectors and potential for breaches.

A new approach to phishing assessments is required to combat these new elements and variables. Instead of a "one size fits all" approach to phishing campaigns, using an intelligent, automated, tiered phishing system that matches the message

difficulty to the user's skill creates a much more realistic approach to simulating the attacks. Organizations deal with threats and attacks every day and from many threat vectors. This study demonstrates the efficacy of such a system and the more granular and relevant metrics it generates. Using tools such as visual analysis and risk scoring or User Behavior Analytics can further enhance the system and make it more useful. Nearly all modern commercial phishing platforms utilize smart or dynamic groups and rules. Thus, for only the investment of time from a practitioner, it is possible to create a tiered phishing system that can help reduce the risk by increasing the skills and abilities of the users to detect and report phishing.

References

Barker, W. C., Scarfone, K., Fisher, W., & Souppaya, M. (2021, September). *Cybersecurity Framework Profile for Ransomware Risk Management*. Retrieved from https://doi.org/10.6028/NIST.IR.8374-draft

Center for Internet Security. (2021, April 4). *The 18 CIS controls* (V 8.0). Retrieved from Center for Internet Security website: https://www.cisecurity.org/controls/cis-controls-list/

Conley, C. (2020, March). *Thoughts on building a tiered phishing system - meeting notes* [pdf].
Sr. Manager, Cybersecurity, Lockheed Martin

Infosec. (2017, July 6). *Security awareness - Definition, history, and types - Infosec resources*. Retrieved October 16, 2021, from https://resources.infosecinstitute.com/topic/security-awareness-definition-history-types/

Joint Task Force. (2021). *SP 800-53 rev. 5, security and privacy controls for info systems and organizations* (800-53). Retrieved from NIST website: https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final

Panaretos, A. (2020, May). *Possibilities and considerations for creating a tiered phishing system - a discussion* [pdf].
Sr. Manager, EY Americas

Cherry, B. (2020, June). *Let's build a tiered phishing system - a discussion* [pdf].
Dir, Information Security Systems, BlueCross BlueShield of Tennessee

Clark, C., Schaff, J. (2020, June) *How do we build a tiered phishing system that's dynamic- a discussion* [pdf].
Principal, Analyst, Business Information Security Systems, BlueCross BlueShield of Tennessee

Parker, G. (2019, June). *Automating response to phish reporting* [Video]. Retrieved from https://www.sans.org/reading-room/whitepapers/email/automating-response-phish-reporting-39000

Parker, G. (2019). *Automating response to phish reporting* (165205). Retrieved from SANS Institute website: https://www.giac.org/paper/gslc/9635/automating-response-phish-reporting/165205

Parker, G. (2021, August). *Building an automated tiered phishing system*. Conference session presented at SANS Security Awareness Summit 2021, Virtual. Retrieved from https://sansorg.egnyte.com/dl/nZVShnn0dh

Parker, G., Cherry, B., & Conley, C. (2020, October). Do you calculate and track the difficulty score for each phishing simulation [Web log post]. Retrieved from https://sth-community.sans.org/category/phishing

Sjouwerman, S. (2019, July). *NEW SANS Whitepaper: Automating response to Phish reporting*. Retrieved from https://blog.knowbe4.com/new-sans-whitepaper-automating-response-to-phish-reporting?hs_preview=xCrdpLUw-11257626783

Smith, J. (2021, September). How is phish-prone percentage calculated?. Retrieved September 10, 2021, from https://support.knowbe4.com/hc/en-us/articles/115010178267

Spitzner, L., & DeBeaubien, D. (2021). *2021 SECURITY AWARENESS REPORT MANAGING HUMAN CYBER RISK*. SANS.org.

Verizon. (2021). *Verizon 2021 Data Breach Investigations Report*. Retrieved from Verizon website: verizon.com/dbir/

Egan, G. (2019). Proofpoint 2019 State of the Phish Report. Retrieved from https://www.proofpoint.com/us/corporate-blog/post/2019-state-phish-report-attackrates-rise-account-compromise-soars

Virustotal.com. (2018). How it works. (n.d.). Retrieved February 4, 2020, from https://support.virustotal.com/hc/en-us/articles/115002126889-How-it-works

Proofpoint. (2019, February 28). Introducing PhishAlarm, Wombat's One-Click Email Reporting Button. Retrieved March 3, 2021, from https://www.proofpoint.com/us/securityawareness/post/introducing-phishalarm-wombats-one-click-email-reporting-button

ITIL Foundation Certification | ITIL. (2019, March 4). Retrieved from https://www.axelos.com/certifications/itil-certifications/itil-foundation-level

Katz, E. (2018, December 7). Phishing Statistics: What Every Business Needs to Know.

Retrieved April 2, 2019, from https://blog.dashlane.com/phishing-statistics/

Anti-Phishing Working Group, Inc. (2019). Technical Whitepapers and Briefings from APWG Sponsors. Retrieved from www.antiphishing.org/resources/technical-whitepapers

Parker, G (2019, January 20). Queries regarding the handling of false positives in phish reporting. Retrieved from https://sth-community.sans.org

Jampen et al. (2020). Don't click: Towards an effective anti-phishing training. A comparative literature review. Retrieved from https://doi.org/10.1186/s13673-020-00237-7

Appendix A

Worksheet

Building an Automated, Tiered Phishing System Planning sheet:

1. How many levels of message templates does my phishing platform have? (Example: 5)
_____

2. How many tiers would I like to use for the users?
a. 3 tiers _____
b. 4 tiers _____
c. 5 tiers _____
d. ? tiers _____

3. How many phishing campaigns a (month/quarter/period) do I want/need to run? _____

4. How many Groups do I need to create?

Percent range 1: _____
Percent range 2: _____
Percent range 3: _____
Percent range 4: _____
Percent range 5: _____
Percent range 6: _____
Percent range 7: _____
Percent range 8: _____
Percent range 9: _____
Percent range 10: _____

5. Sample Rules I think I need:
a. Employee at the company for more than 30 days
b. User Phish Prone Percentage (PPP) is 0%
c. User PPP is more than 1%
d. User PPP is less than or equal to 5%

6. From vendor: a. How do I create/set up smart/dynamic groups?
b. What kind of rules/logic can I apply to smart groups?
c. Who is my CSR (Customer Service Representative) for building tiers?

Geoffrey S. Parker, obiwan324@gmail.com

Appendix B

Hints, Tips, and Best Practices:

- If one creates many tiers, the practitioner may be doing many campaigns - or getting lopsided metrics – unless one groups tiers into campaigns
    - For example, there are three intermediate groups: Create one campaign that phishes all three groups. That represents the intermediate tier
- Balance the tier granularity against the user population and characteristics
    - Most of the users (> 80%) may be in the top 2 tiers:
    - Create one 0% tier
    - Consider creating two advanced tiers and split the percentage of fails up to 10% (1-5% and 6-10%
    - Consider phishing each of these tiers as individual campaigns
- The goal is to get a good, balanced bell curve of the user fails
    - It's okay to move tiers (groups) to different campaigns
    - Also, balance against the levels of difficulty for phishing templates
    - Most platforms have five standard levels of difficulty
- Be careful to increment the PPP when making rules:
    - If tiers overlap (e.g., 1-5%, 5-10% instead of 6-10%), the user will end up in 2 groups! They may be phished twice (2 groups = 2 phishing messages)

- Watch out for decimals (5.9%) – These are not recommended at all! (At some point, multiple users will have the same XX PPP score)
- It could corrupt stats and definitely could lead to disgruntled users.